Effectiveness of an Expert System for Astronaut Assistance on a Sleep Experiment

Gianluca Callini[1], S.M., B.M.E. (1)

Susanne M. Essig, M.S., B.S.A.E. (1)

Dennis M. Heher, M.S., B.A. (2)

Laurence R. Young, Sc.D. (1)

(1) Man-Vehicle Laboratory, Massachusetts Institute of Technology

77 Massachusetts Institute of Technology, Room 37-219

Cambridge, MA 02139

Tel: (617) 253-7805

Fax: (617) 258-8111

(2) Caelum Research Corporation - NASA Ames Research Center

Mail Stop 269-2 (Building 269 Room 235)

Moffett Field, CA 94035

Tel: (650) 604-1084

Fax: (650) 604-3594

Address Manuscript Correspondence to Laurence R. Young, Sc.D.

Running Head: Evaluation of an Expert System

[1] Principal author, MIT Man-Vehicle Laboratory Graduate Research Assistant

ABSTRACT

Background: Principal Investigator-in-a-Box ([PI]) is an expert system designed to train and assist astronauts with the performance of an experiment outside their field of expertise, particularly when contact with the Principal Investigators on the ground is limited or impossible. In the current case, [PI] was designed to assist with the calibration and troubleshooting procedures of the Neurolab Sleep and Respiration Experiment during the pre-sleep period of no ground contact. It displays physiological signals in real time during the pre-sleep instrumentation period, and alerts the astronauts when a poor signal quality is detected. **Methods:** The first (ground based) study presented in this paper required twelve subjects to monitor a set of pre-recorded physiological signals and identify any signal artifacts appearing on the computer screen. Every subject performed the experiment twice, once with the assistance of [PI] and once without. The second part of this study focuses on the post-flight analysis of the data gathered from the Neurolab Mission. After re-playing the physiological signals on the ground, the frequency of correct alerts and false alarms (i.e. incorrect diagnoses by the expert system) was determined in order to assess the robustness and accuracy of the rules. **Conclusions:** Results of the first study indicated a beneficial effect of [PI] and training in reducing anomaly detection time and the number of undetected anomalies. For the in-flight performance, excluding the saturated signals, the expert system had an 84.2% detection accuracy, and the questionnaires filled out by the astronauts showed positive crew reactions to the expert system.

INTRODUCTION

Background

Since it is impractical to allow each Principal Investigator to fly into space with his or her experiment, Principal Investigator-in-a-Box (abbreviated [PI]) was created as an artificial intelligence computer system containing some of the knowledge of a Principal Investigator, to accompany the experiment in space (6).

Rationale for the Study

With the increase in long duration space flights, as well as the increasing time between training and the actual missions anticipated for the International Space Station, expert systems could make a great contribution. A study focused on the evaluation of an expert system with medical diagnostics as its *task domain* (3) would benefit the life sciences community for future long-term space studies. This first study on such a system was divided in two parts: a ground based experiment that used student subjects as "astronaut surrogates", and the actual data gathered from the four science crew members on the Neurolab mission. The evaluation of [PI] for both the ground and flight studies focused on the speed and reliability of the human-computer system in the detection and identification of anomalies in the signals monitored by [PI].

The first version of [PI], also known as the Astronaut Science Advisor (ASA), was used to assist astronauts in the performance of the "Rotating Dome" visual-vestibular interaction experiment on the STS-58 Space Life Sciences 2 (SLS 2) Space Shuttle mission in 1993 (9). This first version of [PI] provided data collection capabilities, as well as protocol assistance, scheduling, and protocol modification suggestions. An additional feature consisted of an "interesting data" filter, designed to perform quick-look data analysis and report any unexpected findings to the astronauts during the experiment.

Extending the successful implementation of the ASA with the Rotating Dome experiment, MIT and NASA Ames Research Center collaborated on the development of a new version of [PI] in conjunction with the "Sleep, Respiration and Melatonin in Microgravity" experiment, created by Dr. Charles Czeisler (Brigham and Women's Hospital, Boston, MA) and Dr. John West (University of California, San Diego). The experiment flew aboard the STS-90 (Neurolab) Space Shuttle mission in April 1998.

Unlike the ASA, however, this new version of [PI] was designed as an integral part of the experiment from the outset. [PI] was designed to assist the Neurolab astronauts with the calibration and troubleshooting of the instrumentation during the pre-sleep period of the experiment when mission rules preclude ground-to-air contact with the crew: the crucial experiment setup and calibration is therefore performed by the crew in isolation in preparation for the extended period of sleep monitoring. [PI]'s role is to display the subjects' physiological signals (divided into cardiorespiratory and electrophysiological signals) during the instrumentation calibration, to identify anomalous signals (alerting the astronauts through a series of signal state lights) and to suggest corrective procedures when necessary (6). The [PI] graphic user interface is shown below (**Fig. 1**).

[Figure 1 Here]

The Sleep experiment is also scheduled for the STS-95 mission to study the effects on sleep of space flight and aging.

The Neurolab Sleep Experiment

A brief overview of the Neurolab Sleep Experiment is necessary in order to fully understand and appreciate [PI]'s function with the Neurolab and STS-95 experiments, as well as the results presented later. There are two goals of the Sleep Experiment. The purpose of the first experiment, devised by Dr. Czeisler, is to study the effect of melatonin on sleep. In order to assess the quality of the astronauts' sleep, it is necessary to monitor several electrophysiological signals: four electroencephalogram (EEG) signals, two electromyogram (EMG) signals and two electro-oculogram (EOG) signals. The second portion of the experiment, developed by Dr. West, studies the effects of microgravity on respiration. This portion requires recording of a series of cardiorespiratory signals: electrocardiogram (EKG), blood oxygen saturation level, PWave (a signal related to the electrocardiogram that is measured with the same device used for the oxygen saturation signal), abdominal and ribcage expansions, nasal airflow and the presence of snoring sounds. The astronauts work in teams of two to apply this instrumentation to each other during the pre-sleep period, a time when communication with the ground support staff is generally discouraged.

The hardware for this experiment consists of the following items:

- Electrode Net (e⊠Net): an elastic web-like cap containing 13 electrode sockets to record the EEG, EMG and EOG signals (Physiometrix, Inc., North Billerica, Massachusetts, U.S.A.)
- "Respiratory Inductance Plethysmography" (RIP) Suit: a lycra tank top and short set containing instrumentation to record abdominal and chest expansions (Blackbottom, Inc. California, U.S.A.);

- "Borg Harness": a bundle of electronic connections and cables for the RIP suit plus instrumentation to measure nasal airflow, EKG, blood oxygen saturation level, PWave and the presence of snoring sounds (manufactured at the NASA Laboratory, The University of California, San Diego, U.S.A.);
- Digital Sleep Recorder (DSR): A device which converts the raw analog signals from the various electrodes and instrumentation to digital signals which are then recorded onto a PCMCIA FlashRAM card (Copyright 1996 Vitaport EDV System GmbH. Distributed by TEMEC instruments BV, The Netherlands).

The flight computer on which [PI] is installed is an IBM ThinkPad laptop equipped with an Intel 486-75 MHz processor and 20 MB of RAM. During the pre-sleep calibration period, the [PI] laptop interfaces with the DSR via an RS-232 serial optical cable. A schematic diagram depicting the manner in which [PI] is connected to the rest of the flight hardware is shown below (**Fig. 2**).

[Figure 2 Here]

The ground-based tests replicated in part the Neurolab Sleep Experiment.

GROUND-BASED EVALUATION[2]

Goal

The experiment that will be presented here was part of a pilot study conducted in January 1998: the experimental protocol was approved by the MIT Committee on the Use of Humans as Experimental Subjects (COUHES). The pilot study was performed to

[2] Large portions of this section are extracted from Gianluca Callini's Unpublished Master's thesis "Assessment of an Expert System for Space Life Sciences: a Preliminary Ground-Based Evaluation of PI-in-a-Box for the Neurolab Sleep and Respiration Experiment," © MIT, all Rights Reserved, September 1998.

acquire preliminary results on the efficacy of [PI]. Student subjects were used to test the hypothesis that an expert system such as [PI] would successfully assist users in the performance of a life sciences experiment relatively outside of their field of expertise. The experiment was conducted following a *comparison-based approach* (4); the results with [PI] were compared to a control condition, with training but no computer decision aid. The goal was also to identify specific aspects of [PI] that influenced subjects' performance during the experiment.

Subjects

Twelve subjects, six male and six female, took part in this experiment. The subjects were all graduate students in the Department of Aeronautics and Astronautics at MIT. The mean age of the subjects is 25 years; only one subject was older than 30 years.

Protocol

The day before the beginning of the experiment, the subjects attended a 1.5 hr. long training lecture. The training introduced the subjects to the identification of electrophysiological sleep data, including the detection of signal anomalies created by improper instrumentation setup or hardware malfunction. The subjects were also introduced to [PI] and its diagnostic capabilities. A live demonstration was given to the subjects by having [PI] play a data file. The experiment was fully described and a short quiz was administered at the end of the session to assess the adequacy of the level of training of the subjects, who mostly received perfect scores.

Experiment Design

The subjects in the evaluation were divided into two groups of six, which began with and without [PI] respectively, and asked individually to monitor a set of pre-recorded electrophysiological signals and to detect and identify each signal artifact displayed on the screen. Due to scheduling and subjects' time constraints, the groups were not balanced by gender. The first group (group A) was composed of four males and two females and the second group (group B) by four females and two males. Acting as his or her own control, every subject performed the experiment with and without the help of [PI]'s diagnostic capabilities on two consecutive days. The groups performed the tests in a crossover fashion as represented in **Table I**:

[Table I Here]

The subjects were provided with a reference manual containing a synopsis of the training session, as well as a list of the anomalies displayed by [PI]. After briefly reviewing the material covered in the training session, the subjects were instructed to start the test session, which lasted about twenty minutes. All twelve subjects completed the experiment and no software or hardware failures were experienced.

The data file the subjects were asked to monitor was real data recorded at the NASA Johnson Space Center during one of the Neurolab crew members' training sessions. The data file contains a total of 59 anomalies for the electrophysiological signals. At least one anomaly appears on every signal displayed. Although the same file was used for all the tests on both days, there were no indications that the subjects acquired enough familiarity with the random appearing of signal artifacts to influence their performances on the second experimental day. [PI] recorded every anomaly onset time and the corresponding subject reaction times.

Results

The average reaction times for the subjects to detect an anomaly, as well as the number of undetected anomalies for both groups, are plotted in **Fig. 3.** Group A ([PI] assistance on Day 1 only) results are on **Fig. 3** (a) and (c), while group B results ([PI] assistance on Day 2 only) are on **Fig. 3** (b) and (d).

[Figure 3 Here]

Members of group B performed the experiment without the assistance of [PI] on the first day and with the assistance of [PI] on the second day. Most of these subjects showed a significant improvement in response time the second day, when [PI] was activated. The average response time for Group B decreased by nearly half on day two with [PI] assistance (**Fig. 3** (b)). Group A, however, which received assistance from [PI] diagnostics on day one, did not show a significant difference in average response time on day two, when [PI] assistance was not given (**Fig. 3** (a)). The average response time decreased only by a minimum amount on day two (without [PI] assistance).

The number of undetected anomalies per subject per day was then analyzed to observe the direct effects and interactions of day and [PI] assistance. It is evident from the graphs that the number of undetected anomalies significantly decreased in group B when [PI] was active on the second day (**Fig. 3** (d)). For the subjects of group A, the number of undetected anomalies was also generally lower when [PI] was active (**Fig. 3** (c)).

Ground Study Discussion

For all the data gathered, an analysis of variance (ANOVA) was performed to determine the significance of several effects on the subject performance. The results obtained from the data analysis presented in this section confirmed the hypothesis that a real-time expert system can positively influence subject performance in the calibration of a space life sciences experiment even with minimal training. Even though the effect of [PI] assistance on the reaction times was not statistically significant by itself, it nonetheless showed a positive influence in improving the overall subject performance. Aside from subject reaction, the most evident effect of [PI] was observed with the number of undetected anomalies, where even the influence of [PI] alone was statistically significant ($p = 0.05$) and improved by 9 anomalies out of a total of 59 presented. The number of undetected anomalies significantly decreased with the help of [PI] regardless of the day that the expert system's assistance was administered. Part of this effect may be attributed to the extremely simple and intuitive appearance of the display.

The analysis of all the reaction times as well as the number of undetected anomalies showed that the cross effect of training and [PI] assistance was also significant ($p = 0.001$ for reaction time and $p = 0.002$ for undetected anomalies). This confirms the importance of the [PI] training session on both the nature of the experiment and the use of the expert system. Even when the effect of day alone was not statistically significant, it still suggested a positive influence. The subjects tended to perform better on the second experimental day, thanks to the experience accumulated on the first day (we cannot, however, ignore the possibility that the re-use of the same test file contributed to the learning effect).

Training is required on *both* the experiment itself and the use of the expert system. There is a danger that the expert system may prove to be counterproductive if the user is not adequately trained to interpret its messages or if the familiarity with the experiment is not satisfactory. This is the reason why the Neurolab crew trained for several months on the sleep experiment as well as the use of the expert systems.

In future ground-based studies, it would be appropriate to greatly increase the number of training hours until the subjects are fully confident with the experimental procedures and the use of the expert system.

FLIGHT PERFORMANCE

Background

The Sleep Experiment was performed on two separate four-day periods during the Neurolab mission. The four science astronauts on the mission were subjects for the sleep experiment and all therefore used [PI] for the pre-sleep calibration and troubleshooting phase of the instrumentation set-up. Post-flight questionnaires were distributed to those who participated in the experiment in an attempt to assess [PI]'s performance and crew interaction during the actual mission. The other set of data is constituted by the first 15 minutes (pre-sleep) of the sleep signals recording. Using [PI] on the ground, it was possible to re-play these signals post-flight and record the anomaly onset times and judge whether [PI]'s heuristics worked correctly, or if false alarms were generated.

Crew Questionnaires

The crew questionnaire results are tabulated in **Table II** below. The questionnaire was composed of yes/no questions and performance ranking questions ranging from 1 (very poor) to 5 (very good).

[Table II Here]

A debriefing of the astronauts after the mission revealed an overall sense of satisfaction about the experiment, including the use of [PI]. In general the responses were positive, with confidence ratings also dependent on the astronauts' background and experience with physiological signal monitoring. The flight performance of [PI] and the feedback from the users also led to several modifications to improve the malfunction correction process.

Data File Analysis

A total of 16 sleep signal recordings were obtained from the Neurolab mission (one per subject per instrumentation session). The files were replayed on the ground to record all the signal artifacts encountered and to assess the accuracy of the [PI] diagnostics. The number of false alarms for each sleep session is tabulated below (**Table III**).

[Table III Here]

For this analysis, the false alarms were determined by two trained data analysts (MIT graduate students working on the [PI] project who gained experience on signal monitoring by working in conjunction with BWH) who re-played the files and examined the data to judge [PI]'s diagnoses for each anomaly. False alarms were defined as cases in which [PI] would alert the astronauts of a poor quality signal when the signal display otherwise showed a good quality signal. In a number of cases, [PI] would activate a red state light for a simple signal saturation, since its rules were not necessarily coded to take that effect into account. These cases were accounted for separately from the false alarms, as the table shows. The astronauts, however, were extensively trained in signal monitoring and were expected to successfully distinguish a saturation signal (possibly due to sudden head movements) from an actual alert. The percentages of valid poor quality identifications were calculated in two different ways, with and without accounting for the saturated signals, which were included in the total number of signals in one case, and ignored in the other case. Omitting the saturation signals increased the percentage of valid diagnoses, since the total number of signals was divided by the sum of the true and false identifications only.

The results in **Table III** show that [PI] performed better on the cardiorespiratory signals than on the electrophysiological signals. This was expected due to the relative simplicity (and robustness) of the cardiorespiratory rules as opposed to the electrophysiological rules. Generally, the rather "noisy" nature of the electrophysiological signals renders the monitoring process more complicated than that for the cardiorespiratory signals. [PI] correctly identified signal artifacts 80.6% of the times on the electrophysiological signals (without counting saturation) and 88.7% of the times on the cardiorespiratory signals. It should be noted that within the two signal categories, certain signal rules were more robust than others. Among the electrophysiological signals, the EEG rules were very accurate (100% correct identifications) as well as the EKG rules (100% correct identifications). The EOG rules, on the other hand, were not as robust and would alert the crew with a red light whenever a signal exceeded the range displayed (instead of waiting for the signal to slowly decay as it normally happens with AC-coupled EOG's). The cardiorespiratory signals which showed the greatest number of problems were the Flow and RIP signals. The [PI] alerts about data problems for the SaO_2 and PWave signals were very accurate and reflected the many problems that the crew reported with the pulse oximeter used to record both readings.

After analyzing the signals individually, a total performance index was calculated by computing the overall percentages of correct signal artifact diagnoses. The results are tabulated in **Table IV** below.

[Table IV Here]

As the table shows, out of all the signal artifacts identified by [PI], 451 were correct diagnoses and 77 were incorrect. Without counting the 100 saturation signals for which [PI] produced a red status light, the system was 84.22% reliable. Counting the saturation warnings, the performance goes down by about 10% but it is still satisfactory. As stated earlier, the saturation signals do not cause a problem if the astronauts are adequately trained in the recognition process and can successfully distinguish a poor signal from a simply saturated one.

The numerical data presented in the table is shown graphically in **Figures 4** and **5**.

[Figure 4 Here]

[Figure 5 Here]

Flight Performance Discussion

The file playback only gives an idea of [PI]'s performance as a monitoring system, but does not provide much information about the amount and type of interaction between the expert system and the astronauts. According to the subjects' comments, they often did not follow the malfunction procedures [PI] displayed but rather corrected the problems by remembering what they had learned in the training sessions. At this stage of the flight performance study, there is no way to find out when the astronauts in fact responded and followed [PI]'s alerts. In order to avoid this kind of inconvenience in the future, [PI] has been updated in preparation for the STS-95 mission, where the sleep experiment will be flown again to study the effects of microgravity on sleep and aging. The new version of [PI] requires the astronauts to click on the state light next to the poor quality signal that the system detected in order to display the corresponding malfunction procedure. During flight, [PI] will record the astronaut mouse click times and the beginning and end of the [PI] activity. The post-flight analysis will allow us to gather data on anomaly onset and end times.

It should also be noted that in several cases, [PI] may not have recognized a poor quality signal which a skilled operator may actually identify. Due to the nature and features of the Neurolab version of [PI], there was no way of accumulating any data on the number of "missed diagnoses" which were instead distinguished by the astronauts. This inability may be compensated for in future versions and application of [PI].

CONCLUSIONS

The Principal Investigator-in-a-Box expert system, designed to aid astronauts with a life sciences experiment outside their field of expertise, has been evaluated. The evaluation was divided in two parts: a preliminary ground-based study involving 12 subjects, and the post-flight analysis from the Neurolab Mission, in which [PI] was used to assist with the Sleep and Respiration in Microgravity Experiment.

The ground-based study revealed a positive effect of [PI] assistance on overall performance (artifact detection time and number of undetected anomalies). A cross effect of [PI] assistance and previous exposure to signal monitoring processes (training) also resulted as a significant factor in subject performance. The flight data confirmed [PI]'s positive effects through the positive feedback from the astronauts, as well as the post-flight data analysis, which showed correct diagnosis percentage of 84.22 % (71.95% counting the saturation signals). The technology should be applicable to the training and "in-flight coaching" aspects of many crew intensive experiments for the International Space Station.

ACKNOWLEDGMENTS

This research was supported by the National Space Biomedical Research Institute, and NASA Ames Research Center, grant number NCC 2-570. The [PI] Team would like to thank Dr. Charles Czeisler, Dr. Derk-Jan Dijk, Dr. James Wyatt, Eymard Riel, Joe Rhonda and Karen Smith of Brigham and Women's Hospital, and Dr. John West, Dr. Kim Prisk, Dr. Ann Elliott and Janelle Fine of the Physiology/NASA Lab at the University of California, San Diego. Many thanks to the entire STS-90 Neurolab crew: Commander Scott Altman, Dr. Jay Buckey, Commander Kathryn Hire, Dr. Richard Linnehan, Dr. James Pawelczyk, Lt. Colonel Richard Searfoss, Dr. Dave Williams, Dr. Alex Dunlap and Dr. Chiaki Mukai, as well as Suzanne McCollum, Sherry Carter, Carlos Reyes and Peter Nystrom of the NASA Johnson Space Center

REFERENCES

1. Callini G. Assessment of an Expert System for Space Life Sciences: a Preliminary Ground-Based Evaluation of PI-in-a-Box for the Neurolab Sleep and Respiration Experiment [Master's Thesis]. Massachusetts Institute of Technology, September 1998.

2. Groleau N. Model-Based Scientific Discovery: A Study in Space Bioengineering [Ph.D. Thesis]. Massachusetts Institute of Technology, September 1992.

3. Feigenbaum E, Friedland P, Johnson BB et al. Knowledge-Based Systems in Japan (Report for the JTEC panel). CACM 1994 May; 37(1): 17-19.

4. Friedman CP, Wyatt JC. Evaluation Methods in Medical Informatics. New York, NY: Springer Inc; 1997.

5. Prerau DS. Developing and Managing Expert Systems – Proven Techniques for Business and Industry. Reading, MA: Addison-Wesley Publishing Company; 1990.

6. Smith RL. Fault Tree Analysis and Diagnostics Development for PI-in-a-Box with the Neurolab Sleep and Respiration Experiment, [Master's Thesis]. Massachusetts Institute of Technology, June 1997.

7. Szolovits P. Uncertainty and Decision in Medical Informatics. Methods Inf Med 1995; 34 (1-2): 111.

8. Young LR. PI-in-a-Box. JISCE 1994 February; 33 (2): 119-22.

9. Young LR, Colombano SP, Haymann-Haber G, et al. An Expert System to Advise Astronauts During Experiments. Proceedings of the International Astronautical Congress; 1989; Malaga, Spain.

TABLES

TABLE I. EXPERIMENTAL TEST MATRIX.

Group	Day 1	Day 2
A	[PI] Diagnostics ON	[PI] Diagnostics OFF
B	[PI] Diagnostics OFF	[PI] Diagnostics ON

TABLE II. POST FLIGHT QUESTIONNAIRE.

Astronaut	How user friendly was [PI]?	Problems starting [PI]?	Did [PI] crash the computer?	Signals that showed more problems than others?	State lights accurate?
1	4	NO	NO	YES (occipital EEG)	YES
2	5	NO	NO	YES (occipital)	"More or Less"
3	4	NO	NO	YES(EOG, PWave, SaO_2)	YES
4	5	NO	NO	YES (occipital)	NO

Astronaut	Confidence on signal display	Confidence on state lights	Did you refer to the paper troubleshooting procedures?
1	5	4	NO
2	5	1	NO
3	5	3	NO
4	5	1	NO

Astronaut	Change to diagnostic messages will simplify operation?	Change to diagnostic messages will speed up operation?	Extent that changes will improve experiment
1	Not great improvement	NO DIFFERENCE	1
2	YES	NA (did not troubleshoot)	1
3	YES	NO DIFFERENCE	3
4	YES	NA (did not troubleshoot)	1

TABLE III. ELECTROPHYSIOLOGICAL AND CARDIORESPIRATORY SIGNAL IDENTIFICATIONS.

Flight Day	Crew ID	TOTAL EP			TOTAL CR		
		True	False	Saturation	True	False	Saturation
3	1	5	0	0	3	2	0
3	3	2	3	10	21	2	0
4	1	5	0	0	5	5	0
4	3	24	6	5	26	2	0
13	3	11	0	15	21	2	0
13	1	11	0	0	24	0	0
15	4	4	4	5	37	3	0
15	2	2	3	10	24	4	0
12	1	15	2	2	27	3	0
12	3	9	5	11	5	0	0
6	4	24	0	10	9	4	0
6	2	24	5	17	14	2	0
14	4	24	11	9	27	1	0
14	2	9	2	5	15	0	0
5	4	6	1	1	18	5	0
5	2	10	10	17	27	0	0
Total		175	42	100	276	35	0
% Valid with Saturation			55.21%			88.75%	
% Valid without Saturation			80.65%			88.75%	

TABLE IV. OVERALL [PI] FLIGHT PERFORMANCE – NUMBER OF SIGNAL ALERTS.

Flight Day	Crew ID	TOTAL ALL			% VALID	
		True	False	Saturation	w/ Sat	w/out Sat
3	1	8	2	0	80.00%	80.00%
3	3	23	5	10	60.53%	82.14%
4	1	10	5	0	66.67%	66.67%
4	3	50	8	5	79.37%	86.21%
13	3	32	2	15	65.31%	94.12%
13	1	35	0	0	100.00%	100.00%
15	4	41	7	5	77.36%	85.42%
15	2	26	7	10	60.47%	78.79%
12	1	42	5	2	85.71%	89.36%
12	3	14	5	11	46.67%	73.68%
6	4	33	4	10	70.21%	89.19%
6	2	38	7	17	61.29%	84.44%
14	4	51	12	9	70.83%	80.95%
14	2	24	2	5	77.42%	92.31%
5	4	24	6	1	77.42%	80.00%
5	2	37	10	17	57.81%	78.72%
Total		451	77	100	71.95%	84.22%

TITLES AND LEGENDS FOR ILLUSTRATIONS

Fig. 1. [PI] Graphic User Interface. The cardiorespiratory window is shown on the top and the electrophysiological window is on the bottom. Note the signal display window, the signal quality state lights and the diagnostics windows on the lower right-hand corner.

Fig. 2. Pre-Sleep Hardware Configuration. The laptop on which [PI] is installed receives the physiological signals via a RS-232 serial cable connected to the DSR. The signals are displayed on the screen in real time.

Fig. 3. Subject Average Reaction Times (a and b) and Number of Undetected Anomalies (c and d) for Groups A and B.

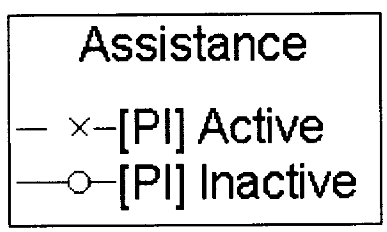


Fig. 4. Electrophysiological (above) & Cardiorespiratory (below) Graphical Representations of the Flight Performance Data. The false alarms, correct identifications and saturation alarms are plotted for every signal. The total performance index is plotted on the far right. Note that no data was charted for the EMG signal because of the absence of poor quality rules in [PI]'s reasoning engine which resulted in no data gathered. Also, there is no real "acceptable" or "poor" quality state for the Microphone signal, therefore [PI] does not perform any quality diagnostics on it, and merely displays it on the signal window.

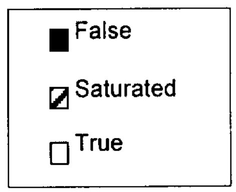
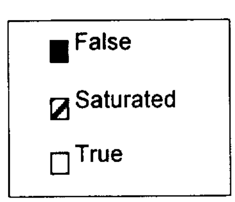
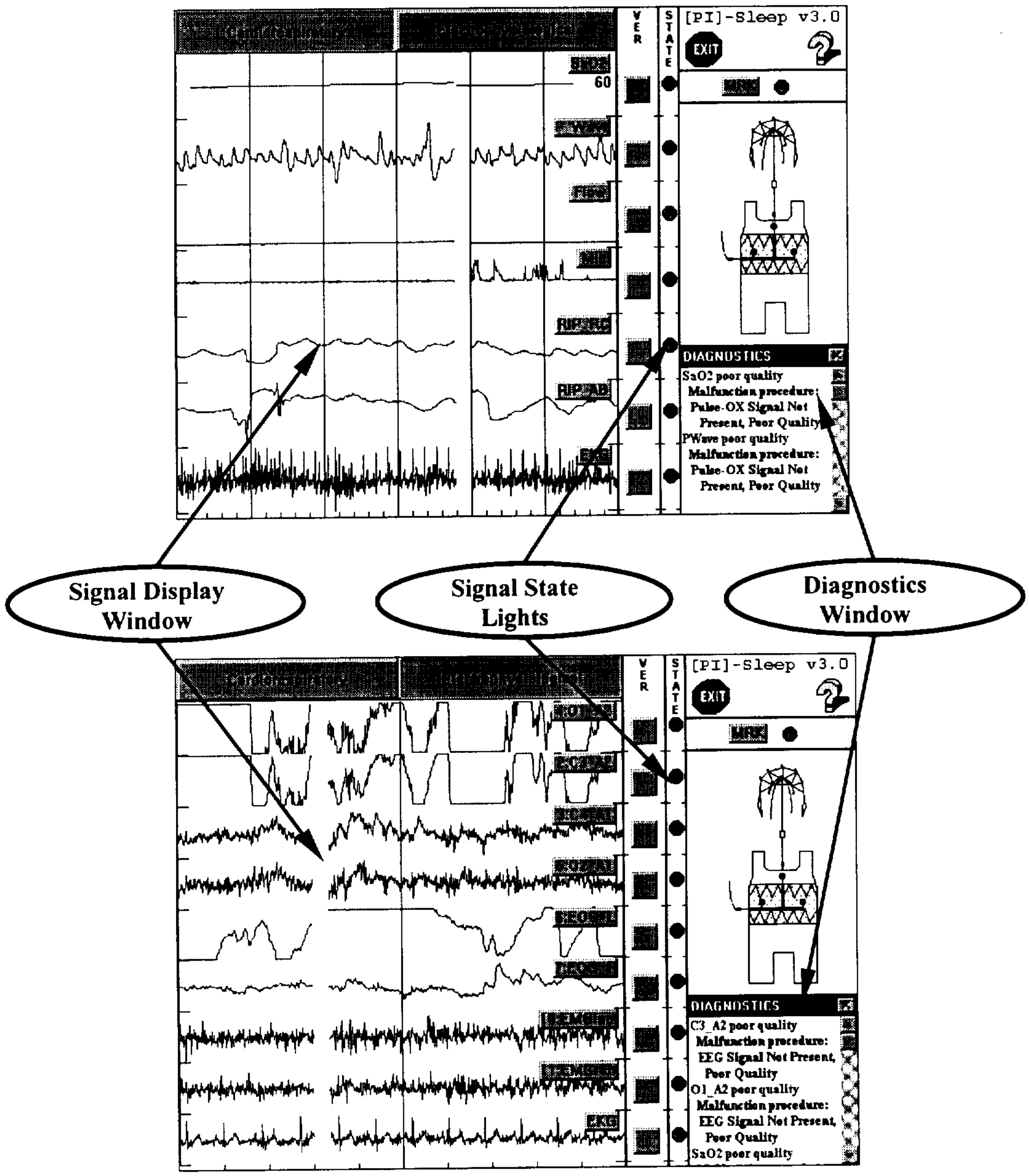
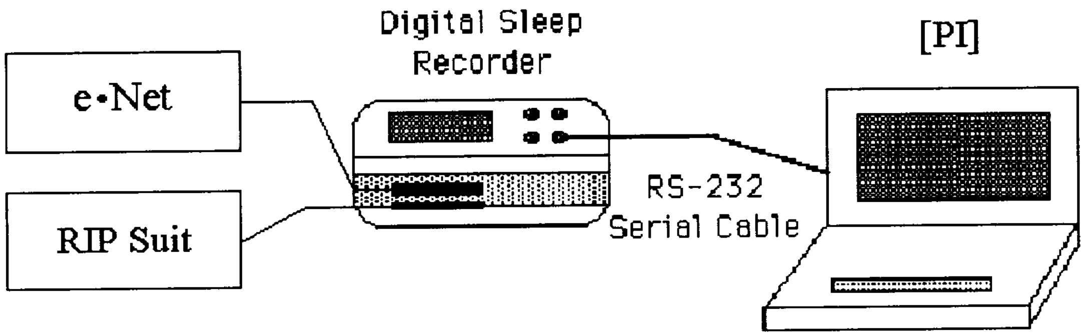


Fig. 5. Final Comprehensive Results of the Flight Performance. As explained above, the Cardiorespiratory (CR) signals were simpler in nature, and were easier to analyze, but presented more anomalies. [PI] diagnosed more saturation alarms and false alarms for the Electrophysiological (EP) signals because of their "noisy" nature, which made them more difficult to analyze for anomalous values. As the far right bars show, [PI]'s performance was positive, with a great number of correct identifications.



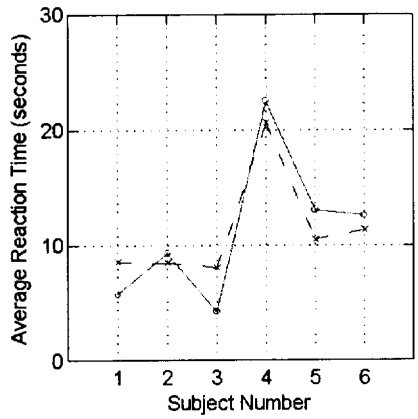
ILLUSTRATIONS


[PI]-Sleep v3.0
EXIT
60
DIAGNOSTICS
SaO2 poor quality
Malfunction procedure:
Pulse-OX Signal Not Present, Poor Quality
PWave poor quality
Malfunction procedure:
Pulse-OX Signal Not Present, Poor Quality
Signal Display Window
Signal State Lights
Diagnostics Window
[PI]-Sleep v3.0
EXIT
DIAGNOSTICS
C3_A2 poor quality
Malfunction procedure:
EEG Signal Not Present, Poor Quality
O1_A2 poor quality
Malfunction procedure:
EEG Signal Not Present, Poor Quality
SaO2 poor quality
EKG

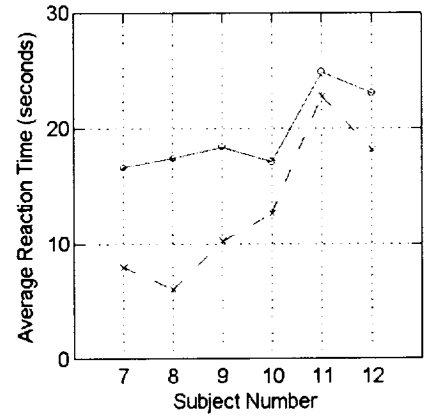

e•Net
RIP Suit
Digital Sleep
Recorder
RS-232
Serial Cable
[PI]

Group A ([PI] on Day 1)

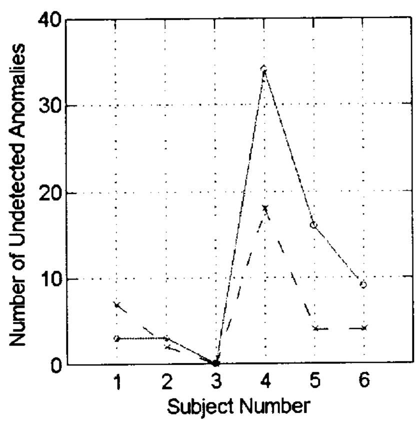
Group B ([PI] on Day 2)



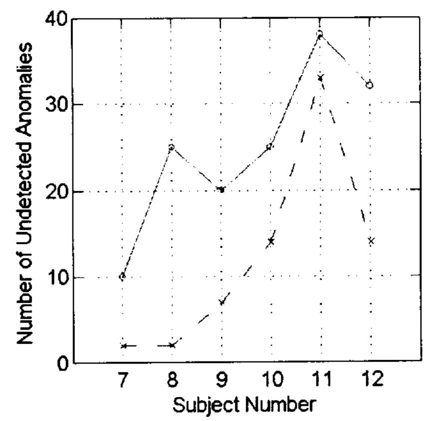
(a)



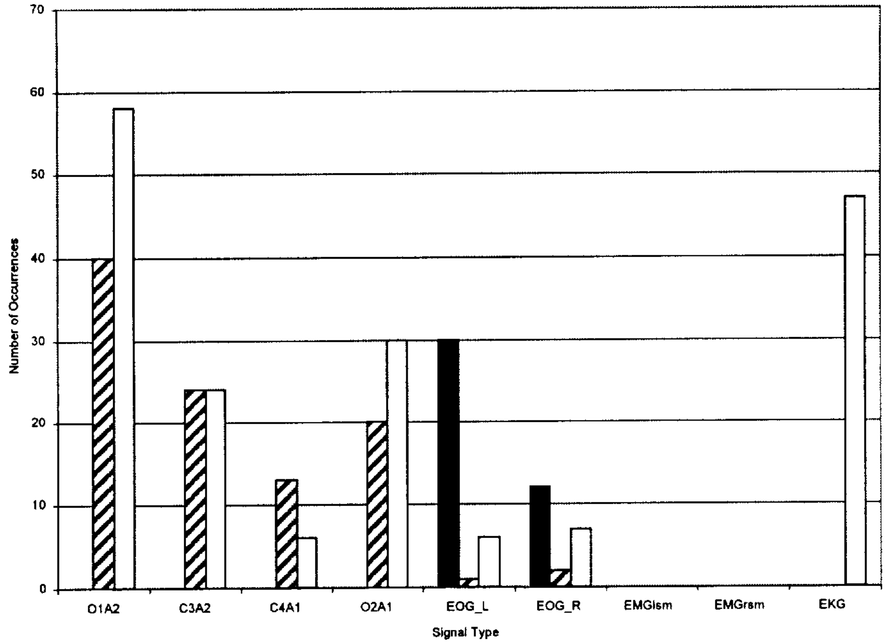
(b)

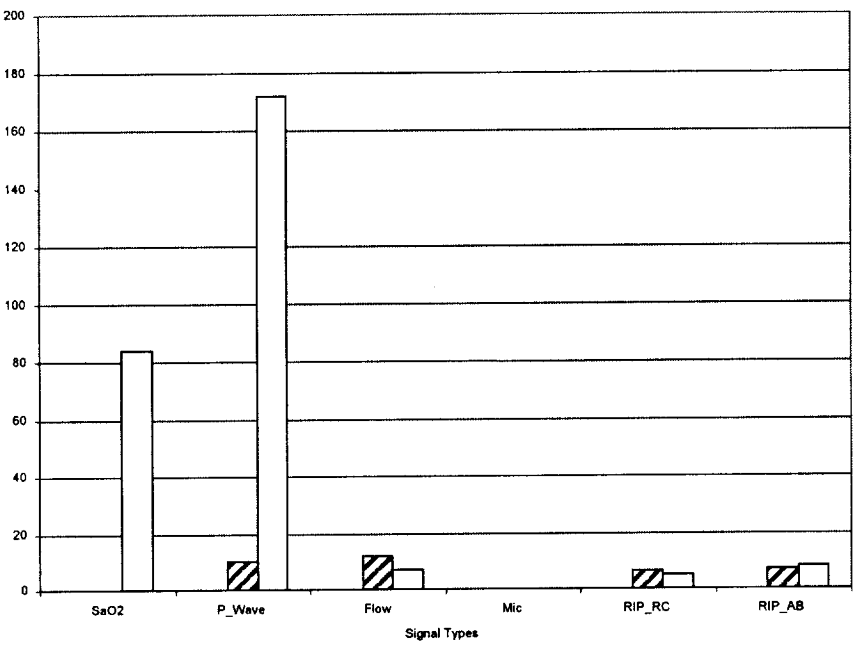


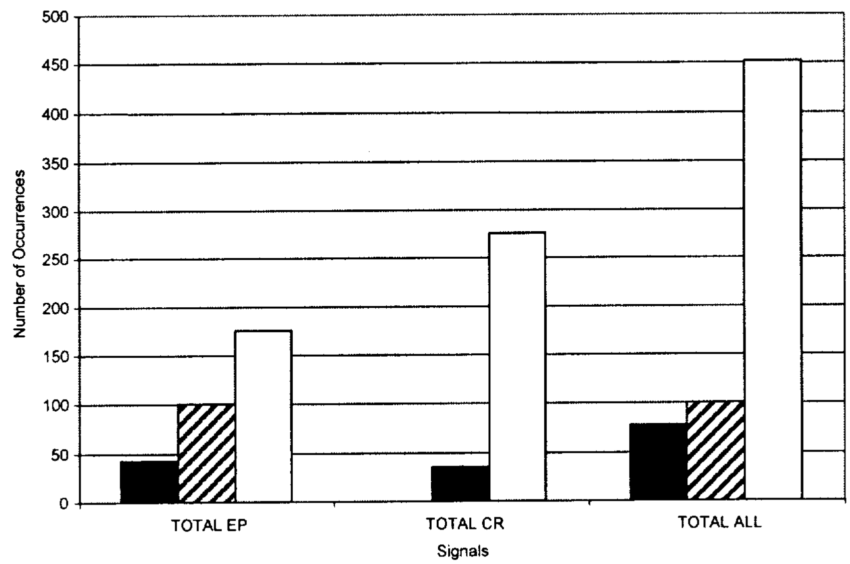
(c)



(d)


Number of Occurrences
70
60
50
40
30
20
10
0
O1A2
C3A2
C4A1
O2A1
EOG_L
EOG_R
EMGlsm
EMGrsm
EKG
Signal Type


200
180
160
140
120
100
80
60
40
20
0
SaO2
P_Wave
Flow
Mic
RIP_RC
RIP_AB
Signal Types


500
450
400
350
300
250
200
150
100
50
0
Number of Occurrences
TOTAL EP
TOTAL CR
TOTAL ALL
Signals